Lecture 23, Wed. 5/18/2022

Kantorovich Duality: Main result, preparations.

Set up: $(X, \mathcal{F})$, $(Y, \mathcal{G})$: measurable spaces.

$$\mathcal{A}(\mu,\nu) = \{\gamma \in \mathcal{P}(X\times Y) : \pi^X_\# \gamma = \mu,\ \pi^Y_\# \gamma = \nu\}$$

$c : X\times Y \to [0,\infty]$: measurable

$\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$.

K-OT: $\inf_{\gamma\in\mathcal{A}(\mu,\nu)} E_K[\gamma]$, $\quad E_K[\gamma] := \int_{X\times Y} c(x,y)\,d\gamma(x,y)$.

Denote

$$\Phi_c = \Phi_c(\mu,\nu) = \{(\varphi,\psi)\in L^1(\mu)\times L^1(\nu) : \varphi(x)+\psi(y) \le c(x,y),$$
$$\text{for } \mu\text{-a.e. } x\in X,\ \nu\text{-a.e. } y \in Y\}.$$

$$J[\varphi,\psi] = \int_X \varphi\,d\mu + \int_Y \psi\,d\nu.$$

The Kantorovich dual problem: $\sup_{(\varphi,\psi)\in\Phi_c(\mu,\nu)} J[\varphi,\psi]$.

Compare this with the discrete OT case (Lecture 5)!

Given $a\in\mathbb{R}^m$, $b\in\mathbb{R}^n$. Set

$$R(C) = \{(f,g)\in\mathbb{R}^m\times\mathbb{R}^n : f\oplus g \le C\},$$
$$f\oplus g = [f_i+g_j] \in \mathbb{R}^{m\times n}.$$

The dual discrete K-OT prob: $\max_{(f,g)\in R(C)} [\langle f,a\rangle + \langle g,b\rangle]$.

Theorem (Kantorovich duality) Let X, Y be Polish, $\mu \in\mathcal{P}(X)$, $\nu\in\mathcal{P}(Y)$, $c: X\times Y\to[0,\infty]$ lower semi-continuous. Then

$$\inf_{\gamma\in\mathcal{A}(\mu,\nu)} E_K[\gamma] = \sup_{(\varphi,\psi)\in\Phi_c(\mu,\nu)} J[\varphi,\psi] = \sup_{(\varphi,\psi)\in\Phi_c(\mu,\nu)\cap(C_b(X)\times C_b(Y))} J[\varphi,\psi].$$

Remarks
- "inf" can be replaced by "min".
- Notation $\Phi_c(\mu,\nu)\cap C_b = \Phi_c(\mu,\nu)\cap\left(C_b(X)\times C_b(Y)\right)$,
$\Phi_c(\mu,\nu)\cap \mathrm{Lip}_b = \Phi_c(\mu,\nu)\cap\left(\mathrm{Lip}_b(X)\times \mathrm{Lip}_b(Y)\right)$.

Recall (Lecture 13)

Proposition Let X and Y be Polish. Let $\gamma\in\mathcal{P}(X\times Y)$.
Then the following are equivalent:

(1) $\gamma\in\mathcal{A}(\mu,\nu)$;

(2) $\gamma(A\times Y)=\mu(A)\ \forall A\in\mathcal{B}(X)$, $\gamma(X\times B)=\nu(B)\ \forall B\in\mathcal{B}(Y)$;

(3) $\int_{X\times Y}\varphi\,d\gamma=\int_X\varphi\,d\mu\ \forall\varphi: X\to\mathbb{R}$: Borel measurable,

$\int_{X\times Y}\psi\,d\gamma=\int_Y\psi\,d\nu$, $\forall\psi: X\to\mathbb{R}$: Borel measurable;

(4) $\int_{X\times Y}\varphi\,d\gamma=\int_X\varphi\,d\mu\ \forall\varphi: X\to\mathbb{R}$: bounded and Borel measurable,

$\int_{X\times Y}\psi\,d\gamma=\int_Y\psi\,d\nu\ \forall\psi: Y\to\mathbb{R}$: bounded and Borel measurable;

(5) $\int_{X\times Y}\varphi\,d\gamma=\int_X\varphi\,d\mu\ \forall\varphi\in C_b(X)$, $\int_{X\times Y}\psi\,d\gamma=\int_Y\psi\,d\nu\ \forall\psi\in C_b(Y)$;

(6) $\int_{X\times Y}\varphi\,d\gamma=\int_X\varphi\,d\mu\ \forall\varphi\in BL(X)$, $\int_{X\times Y}\psi\,d\gamma=\int_Y\psi\,d\nu\ \forall\psi\in BL(Y)$.

Remarks
- $BL(X)=\mathrm{Lip}_b(X)=\{$all bounded and Lipschitz continuous functions on $X\}$.
- The implication (6) $\Rightarrow$ (5) follows from the next proposition.

Proposition Let (X,d) be a Polish space, $\mu\in\mathcal{P}(X)$, and

$\varphi \in L^1(\mu)$. There exist $\varphi_n \in Lip_b(X)$ $(n=1,2,\cdots)$ s.t.

$$\int_X \varphi_n \, d\mu \to \int_X \varphi \, d\mu.$$

<u>Proof</u> Since μ is finite, and $L^1(\mu)$-functions can be approximated by simple functions, we need only to consider $\varphi = \mathbb{1}_A$ for $A \in \mathcal{B}(X)$. By the regularity of μ, $\forall \varepsilon > 0$, $\exists U \subseteq X$, $K \subseteq X$, s.t. U is open, K is compact, $K \subseteq A \subseteq U$, $\mu(U \setminus K) < \varepsilon$. Define $\varphi_\varepsilon(x) = \dfrac{d(x, U^c)}{d(x,K) + d(x,U^c)}$, $x \in X$. Then $\varphi_\varepsilon \in C_b(X)$. In fact $0 \le \varphi_\varepsilon \le 1$. $\varphi_\varepsilon = 1$ on K. $\varphi_\varepsilon = 0$ on U^c. Moreover,

$$\left| \int_X \mathbb{1}_A \, d\mu - \int_X \varphi_\varepsilon \, d\mu \right| = \left| \int_{A \setminus K} d\mu - \int_{U \setminus K} \varphi_\varepsilon \, d\mu \right|$$

$$\le \mu(A \setminus K) + \mu(U \setminus K) \le 2\mu(U \setminus K) = 2\varepsilon.$$

Now, assume $\varphi \in C_b(X)$. First, assume $\varphi \ge 0$. Define

$$\varphi_n(x) = \inf_{y \in X} \{ \varphi(y) \wedge k + k\, d(x,y) \}, \quad x \in X.$$

Then $0 \le \varphi_1 \le \cdots \le \varphi_n \le \cdots \le \varphi \wedge k$, each $\varphi_n \in Lip_b(X)$, and $\varphi_n(x) \to \varphi(x)$ $\forall x \in X$, $\varphi = \sup_n \varphi_n$. See Lecture 13. By the monotone convergence thm, $\int_X \varphi_n \, d\mu \to \int_X \varphi \, d\mu$. In general, $\varphi = \varphi^+ - \varphi^-$, both $\varphi^+, \varphi^- \in C_b(X)$. <u>QED</u>

<u>Proposition</u> Let X, Y be Polish and $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. Then the following are equivalent:

(1) $\gamma \in \mathcal{A}(\mu, \nu)$;

(2) For any $(\varphi, \psi) \in L^1(\mu) \times L^1(\nu)$

$$\int_{X\times Y}[\varphi(x)+\psi(y)]\,d\gamma(x,y)=\int_X\varphi(x)\,d\mu(x)+\int_Y\psi(y)\,d\nu(y);\quad (*)$$

(3) For any $(\varphi,\psi)\in L^\infty(\mu)\times L^\infty(\nu)$, $(*)$ is true;

(4) For any $(\varphi,\psi)\in C_b(X)\times C_b(Y)$, $(*)$ is true;

(5) For any $(\varphi,\psi)\in BL(X)\times BL(Y)$, $(*)$ is true.

Proof. $(1)\Rightarrow(2)$ $\forall\varphi$, $\psi\equiv 1$, $\forall\psi$, $\varphi=1$. $(5)\Rightarrow(4)$: by the previous proposition. The others are clear. QED

Lemma Let $(X,\mathcal{F},\mu)$ and $(Y,\mathcal{G},\nu)$ be two probability measure spaces. Let $c: X\times Y\to[0,\infty]$ be measurable. If $(\varphi,\psi)\in\Phi_c(\mu,\nu)$ and $\gamma\in\mathcal{A}(\mu,\nu)$, then

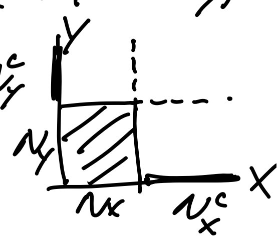
$$\varphi(x)+\psi(y)\le c(x,y),\quad \gamma\text{-a.e. } (x,y)\in X\times Y.$$

Proof Since $(\varphi,\psi)\in\Phi_c(\mu,\nu)$, $\varphi(x)+\psi(y)\le c(x,y)$ for μ-a.e. $x\in X$ and ν-a.e. $y\in Y$. Therefore, there exist $N_X\in\mathcal{F}$, $N_Y\in\mathcal{G}$, s.t. $\mu(N_X)=0$, $\nu(N_Y)=0$, and

$$\varphi(x)+\psi(y)\le c(x,y)\quad \forall (x,y)\in N_X^c\times N_Y^c.$$

Since $\gamma\in\mathcal{A}(\mu,\nu)$, $\gamma(N_X\times Y)=\mu(N_X)=0$ and $\gamma(X\times N_Y)=\nu(N_Y)=0$. Thus, $\gamma((N_X^c\times N_Y^c)^c)=0$, and

$\varphi(x)+\psi(y)\le c(x,y)$ $\forall (x,y)\in N_X^c\times N_Y^c$. QED



Theorem (Weak duality) Let X, Y be Polish, $\mu\in\mathcal{P}(X)$, $\nu\in\mathcal{P}(Y)$, and $c: X\times Y\to[0,\infty]$ lower semi-continuous. Then for any $(\varphi,\psi)\in\Phi_c(\mu,\nu)$ and any $\gamma\in\mathcal{A}(\mu,\nu)$, $J[\varphi,\psi]\le E_K[\gamma]$. Moreover,

$$\sup_{\Phi_c(\mu,\nu)\cap C_b} J\le\sup_{\Phi_c(\mu,\nu)} J\le\inf_{\mathcal{A}(\mu,\nu)} E_K.$$

Proof Let $(\varphi,\psi)\in\underset{\smile}{\Phi_c}(\mu,\nu)$ and $\gamma\in\Pi(\mu,\nu)$. By the above lemma, we have

$$J[\varphi,\psi]=\int_X\varphi\,d\mu+\int_Y\psi\,d\nu=\int_{X\times Y}[\varphi(x)+\psi(y)]\,d\gamma(x,y).$$
$$\le\int_{X\times Y}c(x,y)\,d\gamma(x,y)=E_c[\gamma].\qquad \underline{\text{QED}}$$

Recall the strong duality in linear programing. Given $c\in\mathbb{R}^n$, $b\in\mathbb{R}^m$, $A\in\mathbb{R}^{m\times n}$:

$$\sup_{Ax\le b}c\cdot x=\inf_{y\ge 0,\,A^Ty=c}b\cdot y.$$

The proof uses a minimax principle, related to the method of Lagrangian multipliers, Legendre-transforms, etc.

One idea to prove the Kantorovich's duality (for the continuous OT) is to use the approximation by the discrete duality; cf. e.g., L.C. Evans (1999). Here, we work on the continuous/continuum version of the minimax principle.

Definition (Legendre-Fenchel transform) Let E be a (real) normed vector space and E^* its dual space. The Legendre-Fenchel transform of a function $f:E\to\mathbb{R}\cup\{\infty\}$ is $f^*:E^*\to\mathbb{R}\cup\{\pm\infty\}$:

$$f^*(x^*)=\sup_{x\in E}[\langle x^*,x\rangle-f(x)],$$

where $\langle x^*,x\rangle=x^*(x)$.

Remark If $\mathrm{Dom}(f) := \{x \in E : f(x) < \infty\} \neq \phi$, then $f^* > -\infty$. If f is bounded below by an affine function, then $f^* \not\equiv \infty$. In fact, if $\exists x_0^* \in E^*$ and $c \in \mathbb{R}$ satisfy $f(x) \geq \langle x_0^*, x \rangle + c \ \forall x \in E$. Then, $f^*(x_0^*) \leq -c < \infty$.

Theorem (Fenchel-Rockafellar duality) Let E be a normed vector space and E^* its dual. Let $f, g : E \to \mathbb{R} \cup \{+\infty\}$ be convex. Assume $\exists x_0 \in E$ such that $f(x_0) < \infty$, $g(x_0) < \infty$, and f is continuous at x_0. Then

$$\inf_E (f+g) = \max_{x^* \in E^*} \left[-f^*(-x^*) - g^*(x^*) \right].$$

Proof. Observe

$$-f^*(-x^*) = -\sup_{x \in E} \left[\langle -x^*, x \rangle - f(x) \right]$$

$$= -\sup_{x \in E} \left\{ -\left[\langle x^*, x \rangle + f(x) \right] \right\} = \inf_{x \in E} \left[\langle x^*, x \rangle + f(x) \right].$$

$$-g^*(x^*) = -\sup_{y \in E} \left[\langle x^*, y \rangle - g(y) \right] = \inf_{y \in E} \left[g(y) - \langle x^*, y \rangle \right].$$

Thus,

$$\text{r.h.s. of } (*) = \max_{x^* \in E^*} \inf_{x, y \in E} \left[f(x) + g(y) + \langle x^*, x - y \rangle \right]$$

$$\left[= \inf_{x, y \in E} \sup_{x \in E^*} \left[f(x) + g(y) + \langle x^*, x - y \rangle \right] \right.$$

$\inf_{x,y} \left[f(x) + g(y) + \langle x^*, x - y \right] \leq f(x') + g(y') + \langle x^*, x' - y' \rangle$. Then take $\sup_{x^*}$ and then $\inf_{x', y'}$ to get $\leq$. Also $\forall x^* \in E^*$, $\forall \varepsilon > 0$, $\exists x', y' \in E$ s.t. $\inf_{x,y} \left[f(x) + f(y) + \langle x^*, x - y \rangle \right] \geq$

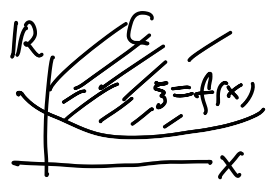
$f(x')+g(y')+\langle x^*, x'-y'\rangle$. Take $\sup\limits_{x^*}$ and $\inf\limits_{x',y'}$.]

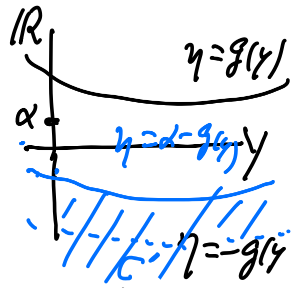
So, need only to show that

$$\sup_{x\in E^*}\ \inf_{x,y\in E}\left[f(x)+g(y)+\langle x^*, x-y\rangle\right]=\inf_{x\in E}\left[f(x)+g(x)\right],$$

with sup being max. By choosing $x=y$, we see that l.h.s $\leq$ r.h.s. So, need only to show that $\exists\, z^*\in E^*$ s.t. $\forall x,y\in E: f(x)+g(y)+\langle z^*, x-y\rangle \geq \inf\limits_E (f+g)=:\alpha$ (*)
If $\alpha=-\infty$ then it's true. $\alpha<\infty$ by the assumption.

So $\alpha\in\mathbb{R}$. Define





$$C=\{(x,\xi)\in E\times\mathbb{R}: \xi>f(x)\},$$
$$C'=\{(y,\eta)\in E\times\mathbb{R}: \eta\leq\alpha-g(y)\},$$

f,g: convex $\Rightarrow$ C, C': convex. Since f is continuous at $x_0\in E$, $(x_0, f(x_0)+1)\in \mathrm{Int}(C)$. So, $\mathrm{Int}(C)\neq\emptyset$, and hence $\overline{C}=\overline{\mathrm{Int}(C)}$. But, $C\cap C'=\emptyset$, since $\alpha=\inf\limits_E(f+g)$. Thus by Hahn-Banach Theorem, $\exists\, l\in (E\times\mathbb{R})^*$, $l\neq 0$, s.t.

$$\inf_{c\in C}\langle l, c\rangle=\inf_{c\in \mathrm{Int}(C)}\langle l,c\rangle\geq\sup_{c'\in C'}\langle l,c'\rangle.$$

Note $l(x,a)=l(x,0)+l(0,a)$ $\forall x\in E, a\in\mathbb{R}$. Let $w^*(x)=l(x,0)$ then $w^*\in E^*$. Also, $\exists\, r\in\mathbb{R}$ s.t. $l(0,a)=ra$ $\forall a\in\mathbb{R}$. Thus, $l=(w^*,r)\neq(0,0)$, and

$$\langle (w^*,r),(x,\xi)\rangle\geq\langle (w^*,r),(y,\eta)\rangle,$$

hence $\langle w^*,x\rangle+r\xi\geq\langle w^*,y\rangle+r\eta$ (*)
if $\xi>f(x)$ and $\eta\leq\alpha-g(y)$. This is possible only if $r>0$: If $r\leq 0$ then for any $x,y\in E$, choose

$\xi > \max(f(x), 0)$, $\eta < \min(\alpha - g(y), 0)$ to get

$$\langle w^*, x\rangle \geq \langle w^*, x\rangle + r\xi \geq \langle w^*, y\rangle + r\eta \geq \langle w^*, y\rangle.$$

Thus $\langle w^*, x-y\rangle \geq 0$. $\forall x, y$. Exchanging x and y, we get $\langle w^*, x-y\rangle \leq 0$. Hence $w^* = 0$. But then by $(*)$ $\xi \leq \eta$ which is impossible as we can choose $\xi \to \infty$ and $\eta \to -\infty$. So, $r > 0$. Now, setting $z^* = w^*/r$, we get $\langle z^*, x\rangle + \xi \geq \langle z^*, y\rangle + \eta$. particularly,

$$\langle z^*, x\rangle + f(x) \geq \langle z^*, y\rangle + \alpha - g(y).$$

for all $x \in X$, $y \in Y$. This leads to $(*)$ QED